OPEN ACCESS

EPIDEMIOLOGICAL SCIENCE

# Genome-wide association of phenotypes based on clustering patterns of hand osteoarthritis identify *WNT9A* as novel osteoarthritis gene

Cindy Germaine Boer ,[1] Michelle S Yau,[2,3] Sarah J Rice,[4] Rodrigo Coutinho de Almeida ,[5] Kathleen Cheung,[4,6] Unnur Styrkarsdottir,[7] Lorraine Southam,[8] Linda Broer,[1] Jeremy Mark Wilkinson ,[9] André G Uitterlinden,[1,10] Eleftheria Zeggini,[8] David Felson ,[11] John Loughlin,[4] Mariel Young,[12] Terence Dante Capellini,[12] Ingrid Meulenbelt,[5] Joyce BJ van Meurs[1]

**Handling editor** Josef S Smolen

For numbered affiliations see end of article.

**Correspondence to**
Dr Joyce BJ van Meurs, Internal Medicine, Erasmus MC, Rotterdam 3000DR, Netherlands;
j.vanmeurs@erasmusmc.nl

Received 1 May 2020
Revised 7 September 2020
Accepted 8 September 2020

## ABSTRACT

**Background** Despite recent advances in the understanding of the genetic architecture of osteoarthritis (OA), only two genetic loci have been identified for OA of the hand, in part explained by the complexity of the different hand joints and heterogeneity of OA pathology.

**Methods** We used data from the Rotterdam Study (RSI, RSII and RSIII) to create three hand OA phenotypes based on clustering patterns of radiographic OA severity to increase power in our modest discovery genome-wide association studies in the RS (n=8700), and sought replication in an independent cohort, the Framingham Heart Study (n=1203). We used multiple approaches that leverage different levels of information and functional data to further investigate the underlying biological mechanisms and candidate genes for replicated loci. We also attempted to replicate known OA loci at other joint sites, including the hips and knees.

**Results** We found two novel genome-wide significant loci for OA in the thumb joints. We identified *WNT9A* as a possible novel causal gene involved in OA pathogenesis. Furthermore, several previously identified genetic loci for OA seem to confer risk for OA across multiple joints: *TGFa*, *RUNX2*, *COL27A1*, *ASTN2*, *IL11* and *GDF5* loci.

**Conclusions** We identified a robust novel genetic locus for hand OA on chromosome 1, of which *WNT9A* is the most likely causal gene. In addition, multiple genetic loci were identified to be associated with OA across multiple joints. Our study confirms the potential for novel insight into the genetic architecture of OA by using biologically meaningful stratified phenotypes.

## Key messages

**What is already known about this subject?**

- Hand osteoarthritis (OA) is one of the most prevalent forms of OA and has a large genetic component, yet only two common genetic loci have been found.
- Lack of findings may be attributed to the modest samples sizes in previous genome-wide association studies (GWAS) and the high disease heterogeneity, which can negatively affect statistical power to robustly identify genetic loci.

**What does this study add?**

- Using three distinct hand OA phenotypes (hand, finger, thumb OA), based on clustering patterns of radiographic OA severity, we have increased the power of our GWAS (n~10 000), and robustly identified a novel genetic locus for thumb OA.
- Functional genomic data from OA disease relevant tissue identified a potential causal variant, predicted to be located in a gene regulatory element, which through chromatin looping interacts with the *WNT9A* promoter to influence *WNT9A* expression.

**How might this impact on clinical practice or future developments?**

- Our results provide the first evidence for WNT9A, a non-canonical Wnt ligand, in human thumb OA.

Check for updates



**To cite:** Boer CG, Yau MS, Rice SJ, *et al*. *Ann Rheum Dis* Epub ahead of print: [*please include* Day Month Year]. doi:10.1136/annrheumdis-2020-217834

## INTRODUCTION

Osteoarthritis (OA) is a serious destructive joint disorder and the third most rapidly rising condition associated with disability.[1] Despite this, no effective treatments that target OA are available. Current treatments only manage pain, not the underlying mechanisms of disease aetiology. An estimated 5% of the world population is affected by OA, with hand OA as one of its most prevalent forms. Given the high prevalence with age and high estimated lifetime risk rates for symptomatic hand OA (39.8%), the number of individuals affected will only continue to increase.[2 3] Hand OA has a high clinical burden, involving considerable pain, deformity and impaired function.[4–6] In addition, hand joints are non-weight bearing, and therefore may reflect the systemic aspects of the disease more than knee and/or hip OA, where mechanical loading is a dominant risk factor.[7] A better understanding of hand OA, its causes and pathophysiological mechanisms is therefore urgently needed.

Hand OA is a complex multifactorial disorder. It shares risk factors, such as repetitive movements

and obesity, with OA at other joint sites.[8 9] Hand OA also has a strong genetic component, with heritability estimates ranging from 39% to 84% depending on the hand joint affected.[10 11] Recently, major advances were made in elucidating the genetic background of hip and knee OA, using large (n>400 000 individuals) genome-wide association studies (GWAS).[12–14] Yet, for hand OA only two common genetic loci have been found near the *MGP* and *ALDH1A2* genes.[15 16] The lack of findings for hand OA may be attributed to the modest samples sizes (n<9000 individuals) in previous GWAS and disease heterogeneity,[17] which is known to reduce power to robustly detect genetic loci.[18]

OA in the hand may be present in any joint, but is most prevalent in the distal interphalangeal joints (DIP), first carpometacarpal (CMC1) and trapezioscaphoid (TS) joints, followed by the proximal interphalangeal joints (PIP). OA is least prevalent in metacarpophalangeal joints (MCP).[19 20] Diverse definitions of hand OA have been used including nodal hand OA (interphalangeal (IP) joints), thumb base OA (CMC1/TS) and generalised hand OA (DIP/PIP/CMC1)[19 20] with varying results, suggesting that disease aetiology may differ between the joints. Moreover, OA affects multiple tissues within the joint including cartilage and bone. This further contributes to disease heterogeneity and warrants the assessment of hand OA by endophenotypes such as joint space narrowing (JSN) and osteophytes (OST) that may capture separate biological processes underlying OA pathology. The use of endophenotypes, quantifiable biological phenotypes intermediate to the genes and the disease, has been successfully used in OA for the detection of novel genetic loci[16 21 22] in knee and hip OA, and may provide new insights into hand OA. For heterogeneous diseases such as OA, stratification of OA phenotypes into different dimensions of disease is one way of increasing power in GWAS.[18]

There are few GWAS of hand OA and none has examined hierarchically defined clusters of OA joint presentation in the hands or hand OA endophenotypes that may provide new insights into hand OA pathogenesis. We therefore set out to examine the occurrence of OA across hand joints and conduct a GWAS stratified by hand OA patterns[23] to identify novel genetic loci for hand OA.

## METHODS

### GWAS, discovery, replication and meta-analysis

We conducted GWAS on radiographic structural phenotypes for OA of the hand using data from the Rotterdam Study (RS) (n∼8700).[24] For a detailed description of the RS, subcohorts and GWAS methods, see online supplemental text and table 1. Briefly, genotypes were imputed to the Haplotype Reference Consortium reference panel (V.1.0) using the Michigan Imputation Server.[25] We assessed genetic associations in each RS sub-cohort using linear regression models adjusted for age, sex and the first four genetic principal components. RVtests[26] was used for the GWAS analyses and results were quality controlled using EasyQC.[27] Variants with an imputation quality <0.3, minor allele frequency <0.05 or effective allele count <5 were excluded and genomic control correction was applied to all SE and p values. Meta-analysis between the discovery cohorts was performed using fixed-effects inverse variance weighting with METAL.[28] Manhattan plots and QQ plots were generated using R and R package qqman.[29] Independent variants with a $p<1\times10^{-6}$ and a $\chi^2$ statistic test of heterogeneity $p>1\times10^{-6}$ were selected for replication in the Framingham Heart Study (FHS) (n=1203).[30] For a detailed description of the FHS, see the online supplemental text. Summary level data from the discovery and replication stage were combined in a joint meta-analysis (METAL).[28] Variants met criteria for replication if the association reached a p<0.05, had the same direction of effect as the discovery sample and reached a joint meta-analysis $p<5\times10^{-8}$. Replicated variants were also examined for association with clinical OA (ie, hospital diagnosed OA) based on GWAS summary statistics from a large-scale OA meta-analysis of data from the UK Biobank and Icelandic deCODE populations.[14 15] For a detailed description of the UK Biobank and deCODE, see the online supplemental text. Associations that reached a p value <0.01 were considered statistically significant.

**Table 1** General characteristics of the study population

| Cohort | RSI | RSII | RSIII |
|---|---|---|---|
| Hand | | | |
| n (OA cases) | 4829 (1830) | 1791 (688) | 2071 (526) |
| Female (%) | 2773 (0.57) | 964 (0.54) | 1180 (0.57) |
| Age (SD) | 67.6 (7.9) | 64.6 (7.9) | 57.1 (7.0) |
| KLsum (aver. (SD)) | 8.4 (9.9) | 6.9 (4.0) | 4.7 (6.5) |
| Osteophytes (aver. (SD)) | 7.1 (7.8) | 6.8 (8.2) | 4.6 (6.3) |
| JSN (aver. (SD)) | 0.84 (2.4) | 0.34 (1.3) | 0.2 (0.8) |
| Finger | | | |
| n (OA cases) | 4839 (1244) | 1803 (474) | 2072 (298) |
| Female (%) | 2779 (0.57) | 972 (0.54) | 1181 (0.57) |
| Age (SD) | 67.6 (7.9) | 64.6 (7.9) | 57.1 (7.0) |
| KLsum (aver. (SD)) | 5.8 (7.4) | 4.5 (6.1) | 3.0 (4.7) |
| Osteophytes (aver. (SD)) | 4.7 (5.5) | 4.3 (5.8) | 2.9 (4.5) |
| JSN (aver. (SD)) | 0.6 (1.9) | 0.25 (1.1) | 0.1 (0.7) |
| Thumb | | | |
| n (OA cases) | 4882 (916) | 1813 (255) | 2083 (166) |
| Female (%) | 2785 (0.57) | 972 (0.54) | 1184 (0.57) |
| Age (SD) | 67.6 (7.9) | 64.6 (7.9) | 57.1 (7.0) |
| KLsum (aver. (SD)) | 2.1 (3.3) | 1.2 (2.3) | 0.8 (1.7) |
| Osteophytes (aver. (SD)) | 1.3 (2.1) | 0.93 (1.7) | 0.6 (1.3) |
| JSN (aver. (SD)) | 0.4 (1.0) | 0.18 (0.63) | 0.1 (0.4) |

aver., average; JSN, joint space narrowing; KL, Kellgren-Lawerence; OA, osteoarthritis; RS, Rotterdam Study.

### Detailed phenotype descriptions

For each participant, all hand joints (16 joints per hand, 32 joints per individual) were scored for Kellgren and Lawrence (KL) grade[31] based on hand radiographs. KL grade is a semi-quantitative score ranging from 0 to 4, where higher scores indicate more severe disease. Radiographic OA was defined as KL grade ≥2 (definite JSN and definite OST). Each joint was also scored for individual radiographic features including JSN and OST.[31] The JSN and OST scores are semi-quantitative scores ranging from 0 to 3, where 0=none, 1=possible, 2=definite and 3=marked.

We conducted hierarchical clustering of KL grade across all hand joints to identify patterns of disease occurrence defined by location and disease severity (see online supplemental text). This yielded three semi-quantitative hand OA phenotypes for analysis: (1) hand KLsum=sum of KL grades across all DIP, PIP, MCP, IP and CMC1 joints in both hands (15 joints per hand, 30 joints per individual, hand KLsum score range: 0–120), (2) finger KLsum=sum of KL grades across all DIP and PIP joints in both hands (8 joints per hand, 16 joints per individual, KLsum score range: 0–64) and (3) thumb KLsum=sum of KL grades across the CMC1 and TS joint (2 joints per hand, 4 joints per individual, KLsum score range: 0–16). Individuals with a missing

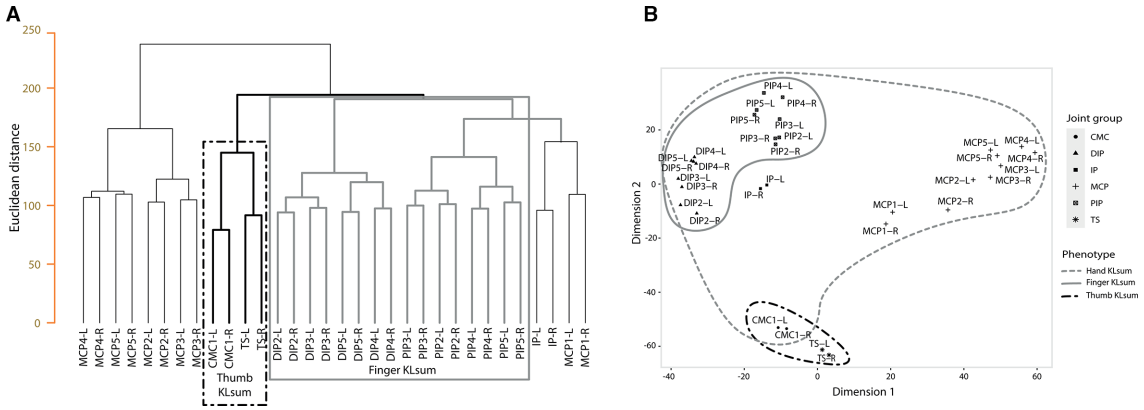



**Figure 1** Tree-dendrogram and multidimensional scaling (MDS) plot of KL scores in the joints of the hand. (A) Tree-dendogram of complete hierarchical clustering of Euclidean distance matrix of KL scores of all hand joints left (L) and right (R). (B) MDS plot of KL scores for all hand joints left (L) and right (R). Data consisted of all RSI, RSII and RSIII (n=8691) participants of whom radiographic X-rays of the hands were made. Selected phenotypes are depicted by the different (dashed) lines. TS, trapezioscaphoid joints; CMC, carpometacarpal joints; DIP, distal interphalangeal joints; IP, interphalangeal joints; KL, Kellgren-Lawrence OA severity score; L/R, left or right joint; MCP, metacarpophalangeal joints; OA, osteoarthritis; PIP, proximal interphalangeal joints; IP, interphalangeal joints; number denotes which joint, ie, PIP2-L, the second PIP joint at the left hand. See online supplemental figures 2 and 3 for tree dendrogram and MDS plots for joint space narrowing and osteophytosis scores.

KL grade in one or more hand joints were excluded from the analysis of phenotypes that required scoring of the missing joint(s). Also, individuals with missing age, sex or genetic principal components were excluded from all analyses (table 1).

### Post-GWAS analysis

Post-GWAS analysis consisted of multiple bioinformatic and functional analyses (online supplemental text). Briefly, all GWAS variants were annotated using HaploReg (V.4.1) and FUMA.[32 33] Intersection of variants with epigenetic markers, proteins, transcription factor (TF) motifs and binding and chromatin interactions was done using data from ROADMAP, ENCODE, HaploReg (V.4.1) and the three-dimensional (3D) genome browser.[34–37] Functional studies included expression quantitative trait loci (eQTL) analysis, methylation expression quantitative trait loci (meQTL) analysis, ATAC-seq analysis and differential gene expression analysis. All functional analyses were performed in human articular cartilage. Details are provided in the online supplemental text.

## RESULTS

### Patterns of osteoarthritis severity in joints of the hand

We used hierarchical cluster analysis on the KL grades of all 32 hand joints of the left hand and right hand to identify clusters (figure 1). Tight symmetric clustering between the left and right joints was seen, in addition to clustering based on joint group, which is in line with previous findings[19 38 39] (figure 1A,B). Clustering based on individual radiographic features, JSN and number of OST produced similar symmetric and joint group clusters (online supplemental figures S1 and 2). We observed consistent clustering of the PIP with the DIP joints and the TS with the CMC joints (figure 1, online supplemental figures S1 and 2). Based on these analyses, we created three semi-quantitative hand OA phenotypes: hand KLsum, thumb KLsum and finger KLsum (table 1).

### Identification of genetic hand osteoarthritis loci

We conducted GWAS on each of the three identified hand OA phenotypes in a discovery sample that included RSI, RSII and RSIII cohorts (n~8700) (online supplemental table S1). In total, we identified seven independent signals with genome-wide suggestive association ($p<1\times10^{-6}$), which were taken forward for replication in the FHS (n=1203) (figure 2 and table 2). In total, four independent signals were genome-wide significant in the meta-analysis ($p\leq5\times10^{-8}$), of which three were significantly replicated ($p<0.05$) (table 2). Two of these signals were novel OA associated loci. The first and most significant novel locus was located on chromosome 1 near the *ZNF678*, *WNT3A and WNT9A* genes, with rs10916199 as the lead single nucleotide variant (SNV). This signal is replicated ($p<0.05$) and genome-wide significantly associated with thumb KLsum (beta=−0.31, $p=2.36\times10^{-13}$). The second replicated novel locus is located on chromosome 11 containing the *F2*, *LRP4* and *CREB3L1* genes, and is associated with thumb KLsum (beta=−0.19, $p=4.7\times10^{-8}$). We also identified two known OA associated loci. The first locus was located near the *MGP* gene, with rs4767133 as the lead SNV. This locus was previously found to be associated with hand KLsum.[16] The second known OA locus, also located on chromosome 12, is the *CCDC91* locus, which was also previously found to be associated with hand KLsum, although it did not reach genome-wide significance.[16] Here, the lead variant rs12049916, was genome-wide significantly associated with hand KLsum (beta=0.78, $p=1.5\times10^{-8}$) and finger KLsum (beta=0.58, $p=2.0\times10^{-8}$), but did not reach nominal significance in the replication cohort, though the direction of effect was the same between discovery and replication cohorts (table 2).

To examine if replicated loci were also associated with clinically defined OA, we looked up findings in GWAS summary statistics from a recent large-scale OA meta-analysis that included the UK Biobank and deCODE populations[14 15] (table 3). Of the novel signals, only rs10916199 (thumb KLsum) was significantly associated with its matching clinical OA phenotype (thumb OA: OR=0.9, 95% CI=0.86 to 0.95, $p=5.7\times10^{-5}$). Of the known signals, rs4764133 was significantly associated with multiple clinical OA phenotypes: thumb OA (OR=1.07, 95% CI=1.03 to 1.12, $p=6.3\times10^{-4}$), finger OA (OR=1.12, 95% CI=1.07 to 1.17, $p=5.7\times10^{-7}$), hand OA (OR=1.09, 95% CI=1.04 to 1.13, $p=6.7\times10^{-5}$) and nominal significantly with knee OA (OR=0.99, 95% CI=0.96 to 1.00, $p=1.8\times10^{-2}$) (table 3). In addition, we also performed a sensitivity analysis in our discovery cohorts to see if the association of rs10916199 with thumb

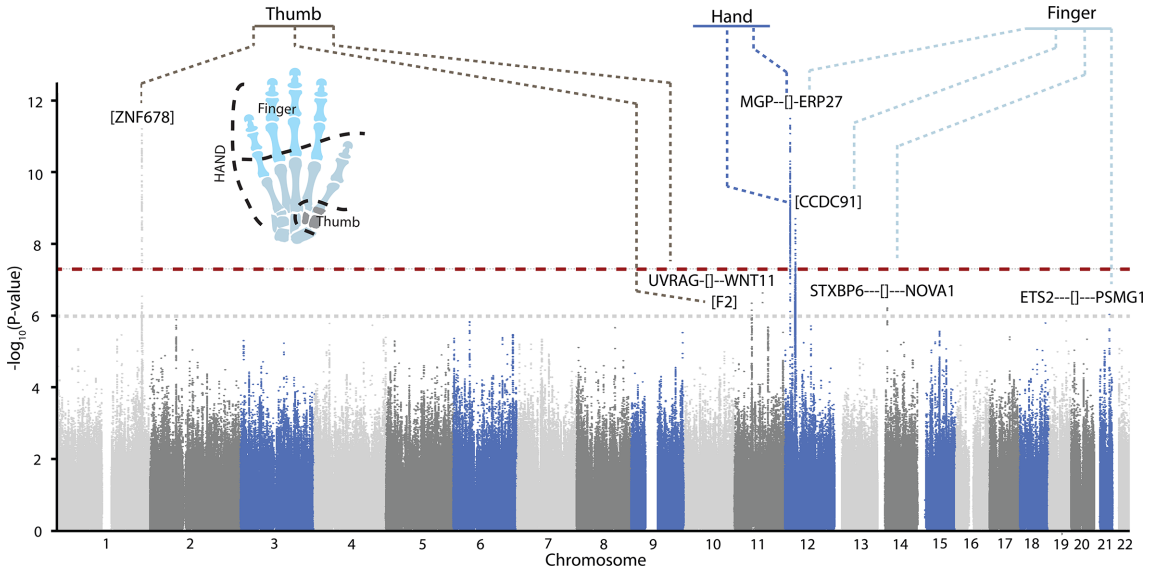



**Figure 2** Combined Manhattan plot of genome-wide association studies (GWAS) discovery results of all radiographic hand OA structural phenotypes. Thumb=thumb KLsum score, finger=finger KLsum score and hand=hand KLsum score. GWAS discovery consists of RSI, RSII and RSIII, and was adjusted for age, sex and the first four genetic principle components. The $-\log_{10}$ p values, for each of the ~11 million single nucleotide polymorphisms (SNPs) analysed (remaining after EASYQC quality control) from the association studies is plotted against their position per chromosome. The dotted red horizontal line corresponds to the genome-wide significant threshold ($p=5\times10^{-8}$). The dotted grey line corresponds to the selection for replication threshold ($p=1\times10^{-6}$). Lead SNP location is represented by [] (intronic), if the SNP is localised intergenic the dashes denotes the distance, - ≤10 kb, -- ≤100 kb, ---- ≤1 Mkb, ---- ≥1 Mkb. For plots of the individual GWAS, see online supplemental figures 3–5.

KLsum was driven by body mass index (BMI. After additional adjustment for BMI, rs10916199 remained strongly associated with thumb KLsum (beta=−0.37, SE=0.05, $p=5.8\times10^{-13}$).

### WNT9A as potential causal gene for thumb OA

We examined the rs10916199 locus in more detail given the strong and consistent association with thumb OA and thumb KLsum. Different levels of information were leveraged for all genes within 1 Mb surrounding rs10916199 in order to prioritise a putative causal gene (figure 3 and online supplemental text). First, we meta-analysed two human osteoarthritic cartilage expression quantitative trait loci (eQTL) datasets (n=116, hip/knee joints, online supplemental table S1)[40 41] and found no significant effect of rs10916199 on gene expression in these datasets. Next, there were significant methylation quantitative trait loci (meQTL) associated with rs10916109 and two CpG sites in human osteoarthritic cartilage (knee/hip joints): CpG09796739 (beta=0.39, FDR $p=8.3\times10^{-7}$) and CpG11520395 (beta=0.21, FDR $p=5.5\times10^{-3}$) (figure 3E, zone 2). These methylation sites are located in a region flanking an active transcriptional start site in primary osteoblasts and chondrogenic cells (figure 3B).

To further assess co-localisation of the identified genetic loci with regulatory function during cartilage differentiation, as many OA loci haven been linked to skeletal development,[42] we intersected the GWAS signals in the locus with accessible chromatin regions (ATAC-seq peaks) in rare human fetal cartilage acquired

**Table 2** Summary of radiographic hand OA structural phenotypes GWAS results

| | | | Discovery* | | | | | | Replication† | | | Meta-analysis‡ | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| rsID | Chr | Pos (hg19) | EA | NEA | EAF | Beta | SE | P value | Beta | SE | P value | Beta | SE | P value | Locus§ |
| Thumb KLsum | | | | | | | | | | | | | | | |
| rs10916199 | 1 | 227 902 472 | A | G | 0.81 | −0.31 | 0.04 | $2.1\times10^{-12}$ | −0.27 | 0.13 | $3.8\times10^{-2}$ | −0.31 | 0.04 | $2.4\times10^{-13}$ | (ZNF678) |
| rs2070852 | 11 | 46 744 925 | C | G | 0.69 | −0.19 | 0.04 | $4.5\times10^{-7}$ | −0.22 | 0.11 | $3.4\times10^{-2}$ | −0.19 | 0.04 | $4.7\times10^{-8}$ | (F2) |
| rs621457 | 11 | 75 858 695 | A | G | 0.52 | −0.18 | 0.03 | $2.4\times10^{-7}$ | −0.07 | 0.11 | $5.2\times10^{-1}$ | −0.17 | 0.03 | $3.3\times10^{-7}$ | UVRAG-[]--WNT11 |
| Finger KLsum | | | | | | | | | | | | | | | |
| rs4764133 | 12 | 15 064 363 | T | C | 0.38 | 0.61 | 0.09 | $5.7\times10^{-12}$ | 1.38 | 0.38 | $2.7\times10^{-4}$ | 0.65 | 0.09 | $4.8\times10^{-14}$ | MGP—[]-ERP27 |
| rs12049916 | 12 | 28 359 985 | G | A | 0.22 | 0.57 | 0.11 | $7.5\times10^{-8}$ | 0.73 | 0.44 | $9.9\times10^{-2}$ | 0.58 | 0.10 | $2.0\times10^{-8}$ | (CCDC91) |
| rs1950427 | 14 | 25 955 502 | T | C | 0.12 | 0.66 | 0.13 | $6.2\times10^{-7}$ | 0.63 | 0.53 | $2.4\times10^{-1}$ | 0.66 | 0.13 | $3.0\times0^{-7}$ | STXBP6 ---[]---NOVA1 |
| rs1029003 | 21 | 40 309 122 | A | G | 0.46 | 0.43 | 0.09 | $9.3\times10^{-7}$ | 0.35 | 0.37 | $3.4\times10^{-1}$ | 0.42 | 0.09 | $5.9\times10^{-7}$ | ETS2---[]---PSMG1 |
| Hand KLsum | | | | | | | | | | | | | | | |
| rs4764133 | 12 | 15 064 363 | T | C | 0.38 | 0.75 | 0.12 | $3.0\times10^{-10}$ | 1.76 | 0.49 | $2.9\times10^{-4}$ | 0.81 | 0.12 | $2.9\times10^{-12}$ | MGP—[]-ERP27 |
| rs12049916 | 12 | 28 359 985 | A | G | 0.77 | 0.77 | 0.14 | $5.3\times10^{-8}$ | 0.90 | 0.56 | 0.11 | 0.78 | 0.14 | $1.5\times10^{-8}$ | (CCDC91) |

*Discovery consists of RSI, RSII, RSIII, samples sizes per phenotype are: thumb KLsum n=8778; finger KLsum n=8714; hand KLsum n=8629.
†Replication cohorts consist of Framingham Heart Study n=1203.
‡Meta-analysis is discovery and replication cohorts using inverse variance weighted meta-analysis using METAL.
§SNP location represented by [], if the SNP is localised intergenic the dashes denotes the distance, - ≤10 kb, -- ≤100 kb, --- ≤1 Mkb, ----- >1 Mkb.
EA, effect allele; EAF, effect allele frequency; GWAS, genome-wide association studies; NEA, non-effect allele; RS, Rotterdam Study; SNP, single nucleotide polymorphism.

**Table 3** Association with hip, knee, hand, finger and thumb OA

| rsID | EA | EAF | Thumb OA (7280 cases; 605 132 controls) | | | Finger OA (7037 cases; 222 772 controls) | | | Hand OA (8591 cases; 224 326 controls) | | | Hip OA (17 151 cases, 613 790) | | | Knee OA (24 919 cases, 613 702 controls | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | OR | 95% CI | P value | OR | 95% CI | P value | OR | 95% CI | P value | OR | 95% CI | P value | OR | 95% CI | P value |
| rs10916199 | A | 0.81 | 0.91 | 0.86 to 0.95 | $5.7\times10^{-5}$ | 0.98 | 0.93 to 1.04 | 0.57 | 1.03 | 0.98 to 1.08 | 0.32 | 1.00 | 0.98 to 1.03 | 0.95 | 0.99 | 0.97 to 1.02 | 0.59 |
| rs2070852 | C | 0.69 | 0.97 | 0.89 to 1.05 | 0.47 | 0.99 | 0.94 to 103 | 0.58 | 0.99 | 0.96 to 103 | 0.80 | 0.98 | 0.96 to 1.01 | 0.22 | 0.99 | 0.97 to 1.01 | 0.40 |
| rs4764133 | T | 0.39 | 1.07 | 1.03 to 1.12 | $6.3\times10^{-4}$ | 1.12 | 1.07 to 1.17 | $5.7\times10^{-7}$ | 1.09 | 1.04 to 1.13 | $6.7\times10^{-4}$ | 0.99 | 0.96 to 1.01 | 0.30 | 0.98 | 0.96 to 1.00 | 0.018 |
| rs12049916 | A | 0.77 | 1.00 | 0.96 to 1.05 | 0.87 | 0.99 | 0.94 to 1.04 | 0.67 | 0.99 | 0.94 to 1.04 | 0.68 | 0.96 | 0.93 to 0.99 | $2.9\times10^{-3}$ | 0.99 | 0.97 to 1.01 | 0.27 |

Thumb OA, finger OA, hand OA GWAS summary statistics are from the deCODE cohort. Hip OA and knee OA summary statistics are from a meta-analysis of deCODE and UK Biobank cohorts.
EA, effect allele; EAF, effect allele frequency; NEA, non-effect allele; OA, osteoarthritis.

from proximal and distal long bones from gestational day(E) 59 of development.[43] Two of the SNVs in high LD ($r^2 \geq 0.8$, figure 3A) with rs10916199 (rs74140304 and rs11588850) intersected with accessible chromatin regions across multiple human long bone cartilage (figure 3C and 3D). In addition, rs74140304 also intersected with an active transcription start site in osteoblast and chondrogenic cells (figure 3B).[28] Next, we examined 3D chromatin conformation in the locus. Since no chromatin conformation capture data were available for chondrogenic or bone cells, we examined data from human mesenchymal stem cells, which are stem cell progenitors for chondrocytes and osteoblasts (figure 3G).[28] The genomic location of rs1158850 (figure 3G, zone 2) appears to come into close proximity with the promoter region of *WNT9A* (figure 3G, zone 3). In addition, CTCF binding peaks in osteoblasts also intersect with the *WNT9A* promoter region (figure 3F, zone 3), and are located near rs11588850 (figure 3F, zone 2).

Next, differential expression analysis between OA lesioned and preserved cartilage (hip/knee joints) identified *WNT9A* as the most significant result, with increased expression in OA lesioned cartilage (n=21, fold change=2.42, $p=9.4\times10^{-8}$, online supplemental table S2). Lastly, we examined whether rs11588850 may significantly affect ($p<4.0\times10^{-8}$) the regulatory TF binding motifs[37] located within the *WNT9A* promoter.[35] The minor allele (G) of rs11588850 is in high linkage disequilibrium ($r^2 > 0.8$) with the OA risk increasing allele (G) of rs10916199, which significantly increases the binding affinity of the TF binding motif for RAD21 (G-allele logarithm of odds (LOD)=11.2, A-allele LOD=9.8). The RAD21 protein has been previously shown to bind to the *WNT9A* promoter region (online supplemental table S3). Thus, our results indicate *WNT9A* as novel OA associated gene, where rs1158850 is a potential regulatory variant for *WNT9A* (figure 3, zones 2–3).

### Additional hand osteoarthritis associated loci

OA is highly heritable and co-occurrence of OA in multiple joint sites is well recognised[44] . As the hand joints are non-weight bearing, causes of OA in these joints may reflect effects of systemic risk factors, unlike the hip and knee joint where mechanical loading is a dominant risk factor.[7] Thus, we examined whether other known OA loci may also confer risk for hand OA (figure 4). For 29 of the previously reported OA associated SNVs,[13 14] nominally significant associations ($p<0.05$) were observed for one or more hand OA phenotypes. Strong associations were seen for known hand OA loci: *ALDH1A2-locus* (rs3204689), *MGP-locus* (rs4767133) (figure 4A,B) and *COG5* (rs3815148),[45] an SNV identified from a candidate gene study for hand OA. Interestingly, the *MGP*-loci and *ALDH1A2*-loci, were only associated with finger and/or hand OA phenotypes, but not with thumb OA. In contrast, the *BCL7A*-locus (rs11059094), was only associated with thumb OA. Strikingly, several reported knee and hip OA loci were also associated with hand OA phenotypes in our study (Bonferroni, $p<5.8\times10^{-4}$): *RUNX2-locus* (rs12154055), *COL27A1-locus* (rs919642), *ASTN2-locus* (rs13253416), *IL11-locus* (rs4252548), *TGFa-locus (rs3771501)* and *GDF5-locus* (rs143384) (figure 4B). Since some of these known loci were previously found to be associated with knee and/or hip OA, they may reflect common mechanisms across all joints in OA.

## DISCUSSION

We identified four genome-wide significant loci associated with hand OA phenotypes, of which two were novel and specific for thumb OA. Integration of multiple lines of data provided

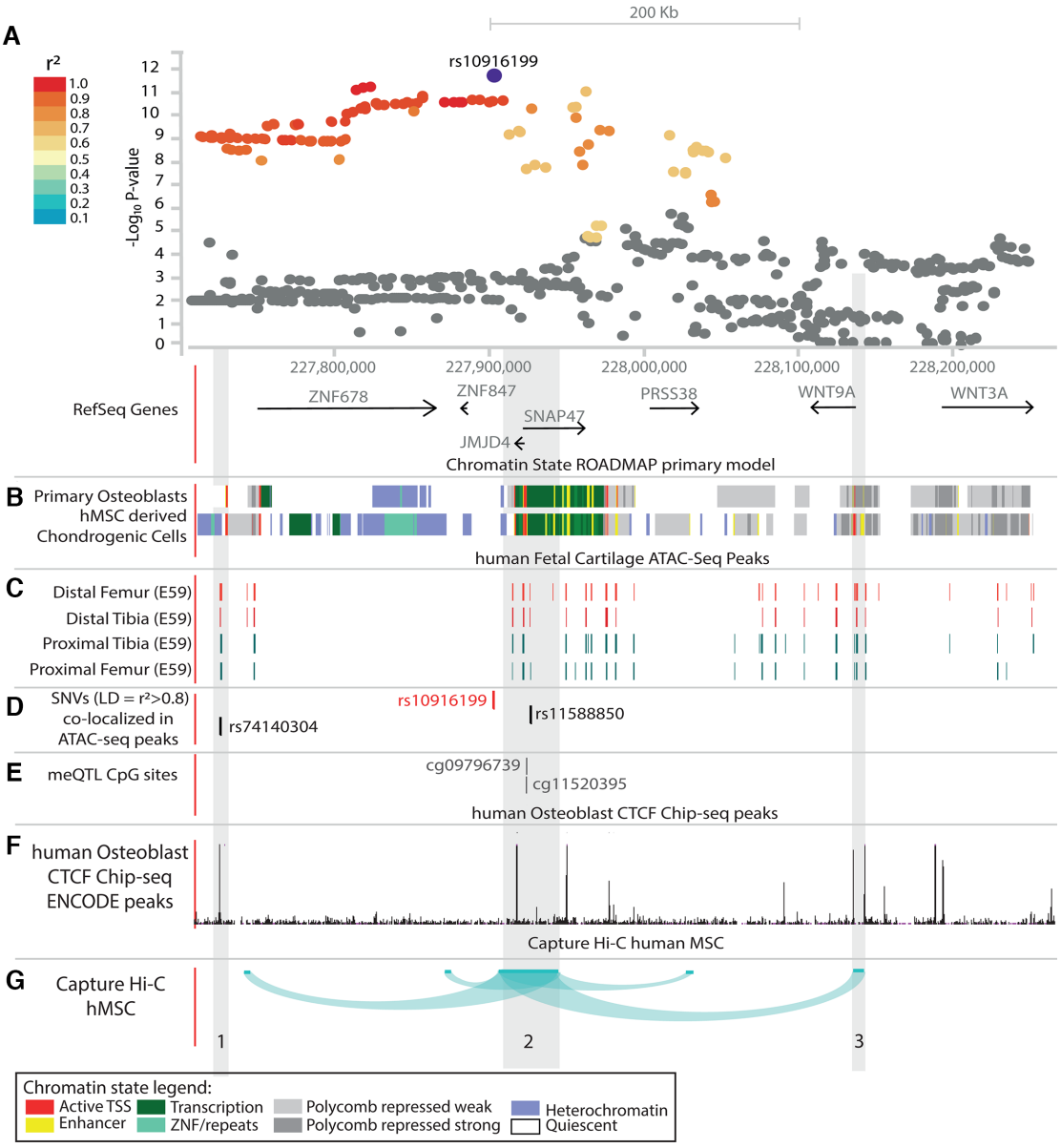



**Figure 3** Schematic overview of part of the rs10916199 locus. (A) LocusZoom plot of rs10916199 locus, the Y-axis depicts the $-\log_{10}$ p value of the single nucleotide variant (SNV) from the thumb Kellgren-Lawrence (KL)sum genome-wide association studies (GWAS). Colours depict the linkage disequilibrium (LD) ($r^2$) between the variant and the LD SNV rs10916199. The X-axis depicts the relative genomic location, depict are the protein coding genes at those genomic locations. For this genomic region depicted in figure (B–G) are several epigenetic annotations are plotted. (B) Chromatin state, as predicted by the ROADMAP 15-state model on histone modifications, for primary osteoblasts and human mesenchymal stem cell-derived chondrocytes. Colours depict chromatin states, legend at bottom of full figure. (C). ATAC-seq peaks from human embryonic cartilage at different bone development sites at gestation day(E) 59. (D) Location of the lead SNV, rs10916199 and two putative causal SNVs which co-locative with ATAC-seq peaks. (E) Genomic location of rs10916199 meQTL CpG sites. (F) Chip-seq CTCF protein binding peaks from human primary osteoblasts from ENCODE. (G) Capture Hi-C chromatin interactions from three-dimensional (3D) genome browser for human mesenchymal stem cells (hMSC) and mesoderm. Depicted are the chromatin interactions from the promoters of the queried gene to the genomic location of interaction, this was done for the JMJD4/SNAP47 transcription start site (TSS), WNT9A TSS and the WNT3A TSS. For details on methods and underlying data, see online supplemental methods.

cumulative evidence that *WNT9A* may be a causal gene for thumb OA. We first conducted a cluster analysis of hand joints to identify less heterogeneous clusters of joints that served as the basis of the hand OA phenotypes assessed in this GWAS. With this approach of using radiographically defined biologically relevant OA phenotypes to reduce phenotype heterogeneity and increase statistical power, we were able to robustly identify known and novel OA loci despite our modest sample size (n~9900).[18 22] This indicates that assessment of stratified phenotypes in OA may be warranted to improve GWAS statistical power and provide novel insight into OA biology.

Using bioinformatic analysis and functional genomics datasets, we were able to identify rs1158850 as potential causal variant. This SNV is nearby meQTL CpGs and the G allele of this variant is predicted to increase RAD21 (RAD21 Cohesin Complex Component) binding affinity in a region that has chromatin interactions with the *WNT9A* promoter. RAD21 is a part of the cohesion complex, involved in the formation of chromatin

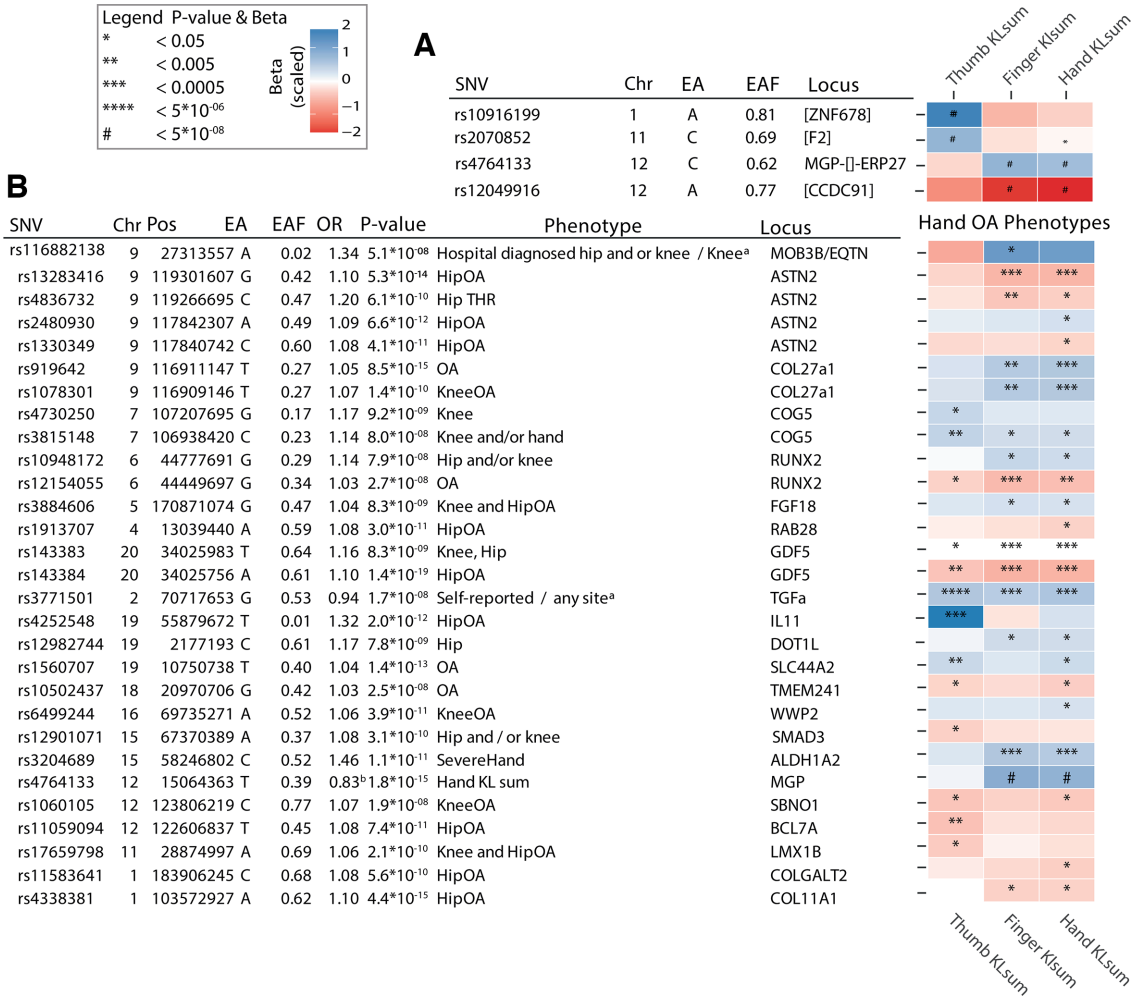



**Figure 4** Heatmap depicting the effect of osteoarthritis (OA) associated single nucleotide variants (SNVs) in each hand OA phenotype. (A) The found associated lead SNVs of the hand OA phenotypes and their beta and p value in the other phenotypes. (B) Depicted are all known OA SNVs which had a nominal significant effect in at least one stratified hand OA phenotype. All betas were calculated for the reported effect allele and scaled. Colours represent the scaled beta of the effect allele, which is here the minor allele. P values are represented by * in the box. Chr, chromosome; EA, effect allele; EAF, effect allele frequency; Gene, reported gene from the GWAs study; KL, Kellgren-Lawrence score; Pos, base pair position on the chromosome Hg19.

loops with CTCF.[46] Both RAD21 and CTCF bind to the *WNT9A* promoter region. In line with these findings, *WNT9A* expression was significantly increased in OA lesioned cartilage compared with preserved OA cartilage. Combining all results, we postulate that rs1158850 is located in a gene regulatory element, increases RAD21 binding, and mediated by CTCF, interacts with the *WNT9A* promoter to influence *WNT9A* expression.

*WNT9A* (wingless-type MMTV integration site family, member 9A) previously known as *WNT14*, is a member of the *WNT* gene family, and has been shown to play a central role in synovial joint formation.[47 48] Knockout *WNT9A* mice have severe skeletal developmental defects, and are neonatal lethal.[49] Expression of *WNT* members by chondrocytes leads to the destruction of the cartilage matrix by the upregulation of Wnt/β-catenin signalling. Inhibition of WNT members has been suggested as a plausible OA therapeutic strategy, with recent success in a murine model of OA.[50 51] However, this is the first evidence for WNT9A, a non-canonical Wnt ligand, in human OA.

In addition, to identifying novel OA loci, we also provide evidence for a subset of generalised OA genetic risk loci: *GDF5* (rs143384), *TGFα* (rs2862851/rs3771501), *RUNX2* (rs12154055), *ASTN2* (rs2480930, rs13823416), *COL27A1* (rs919642) and *IL11* (rs4252548). These loci should be given priority as potential therapeutic targets since genetically supported drug targets have been shown to double the success rate of therapeutics in clinical development and intervention at these target loci may be beneficial regardless of which joint site is affected by OA.[52]

Although collectively our findings implicate *WNT9A* in thumb OA, there are several limitations. First, age is the most predominant risk factor for OA, yet the genetic background may determine the age of onset, rather than the lifetime risk for OA. Therefore, future genetic studies may benefit from examining the age of onset rather than adjust for age.[53] Second, our functional findings are based on chondrogenic data sourced from several different tissues that did not include tissues from the hand joints. However, consistent results were found across the available chondrogenic source material from several different origins (primary, cell culture, developmental), indicating a more general role for the *WNT9A* locus in chondrocyte functional pathways. Given the complex nature of OA susceptibility and the fact that pathophysiological causes are not uniform across skeletal sites, alterations in *WNT9A* expression may be seen in other

joints, but may have a more marked detrimental effect on the thumb joints. The lack of genetic association of WNT9A variants for knee and hip OA might be due to the phenotype definition of these GWAS studies: self-reported OA and 10th revision of the International Statistical Classification of Diseases codes, which do not necessarily correspond or have enough statistical power to identify genetic variants associated to radiographic phenotypes.

In summary, by examining the distribution of radiographic OA features in the hand joints, we identified three distinct hand OA phenotypes that provided the basis for identification of a novel locus for thumb OA despite our modest sample size. We identified *WNT9A* as a plausible causal gene for thumb OA, providing new insights into the genetic architecture of hand OA and a new candidate for OA therapeutic development.

**Author affiliations**
[1]Department of Internal Medicine, Genetic Laboratories, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[2]Hebrew SeniorLife, Beth Israel Deaconess Medical Center. Harvard Medical School, Hinda and Arthur Marcus Institute for Aging Research, Boston, Massachusetts, USA
[3]Department of Rheumatology, Boston University School of Medicine, Boston, Massachusetts, USA
[4]Biosciences Institute, Newcastle University, Newcastle upon Tyne, UK
[5]Department of Biomedical Data Sciences, Section Molecular Epidemiology, Leiden University Medical Center, Leiden, The Netherlands
[6]Newcastle University, Bioinformatics Support Unit, Newcastle upon Tyne, UK
[7]deCODE genetics/Amgen, Inc, Reykjavik, Iceland
[8]Institute of Translational Genomics, Helmholtz Zentrum München Deutsches Forschungszentrum für Gesundheit und Umwelt, Neuherberg, Germany
[9]Department of Oncology and Metabolism, University of Sheffield, Sheffield, UK
[10]Department of Epidemiology, Erasmus MC, University Medical Center, Rotterdam, The Netherlands
[11]Arthritis Research UK Epidemiology Unit, The University of Manchester, Manchester, UK
[12]Human Evolutionary Biology, Harvard University, Cambridge, Massachusetts, USA

**Twitter** Cindy Germaine Boer @CurlyGeneticist and Terence Dante Capellini @TDCapellini

**Acknowledgements** The authors would like to thank the study participants, the staff from the Rotterdam Study and the participating general practitioners and pharmacists. The generation and management of GWAS genotype data for the Rotterdam Study (RSI, RSII, RSIII) was executed by the Human Genotyping Facility of the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, Rotterdam, The Netherlands. The authors would like to thank Pascal Arp, Mila Jhamai, Marijn Verkerk, Lizbeth Herrera and Marjolein Peters, MSc, and Carolina Medina-Gomez, MSc, for their help in creating the GWAS database, and Linda Broer PhD, for the creation of the imputed data.

**Contributors** CGB designed the study, performed the analyses, made the figures and tables and wrote the manuscript. MSY performed the replication analysis. SJR, RCdA, KC and LS performed functional analyses and look-ups (eQTL, meQTL and differential expression), MY and TDC performed ATAC-seq analysis, US provided GWAS data of deCODE and UK Biobank, LB provided analysis help. AGU provided access to the Rotterdam study dataset, DF provided replication data, EZ, JL, TDC and IM provided functional data. JBJvM designed the study and supervised this work. All authors critically assessed the manuscript.

**Funding** The Rotterdam Study is funded by Erasmus Medical Center and Erasmus University, Rotterdam, Netherlands Organization for the Health Research and Development (ZonMw), the Research Institute for Diseases in the Elderly (RIDE), the Ministry of Education, Culture and Science, the Ministry for Health, Welfare and Sports, the European Commission (DG XII) and the Municipality of Rotterdam. The Rotterdam Study GWAS datasets are supported by the Netherlands Organisation of Scientific Research NWO Investments (nr. 175.010.2005.011, 911-03-012), the Genetic Laboratory of the Department of Internal Medicine, Erasmus MC, the Research Institute for Diseases in the Elderly (014-93-015; RIDE2), the Netherlands Genomics Initiative (NGI)/Netherlands Organisation for Scientific Research (NWO) Netherlands Consortium for Healthy Aging (NCHA), project nr. 050-060-810. The Framingham Heart Study of the National Heart, Lung and Blood Institute of the National Institutes of Health and Boston University School of Medicine was supported by the National Institutes of Health (contract no. HHSN268201500001I, N01-HC-25195, AG18393, AR47785) and its contract with Affymetrix, Inc. for genotyping services (N02-HL-6-4278). Analyses reflect intellectual input and resource development from the Framingham Heart Study investigators participating in the SNP Health Association Resource (SHARe) project. MSY was supported by the National Institute of Arthritis and Musculoskeletal and Skin Diseases (NIAMS) and the National Institute on Aging (NIA) (R01AR075356). Rice, Cheung and Loughlin were supported by vs Arthritis (grant 20771), by the Medical Research Council and Arthritis Research UK as part of the MRC-Arthritis Research UK Centre for Integrated Research into Musculoskeletal Ageing (CIMA, grant references JXR 10641, MR/P020941/1 and MR/R502182/1) and by the European Union's Seventh Framework Programme for research, technological development and demonstration under grant agreement number no. 305 815 (D-BOARD). The research leading to the RAAK biobank and the current results has received funding from the Dutch Arthritis Association (DAA 2010_017) and the European Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement no. 259 679. ATAC-seq datasets generated by MY and TDC were funded by a Harvard University Dean's Competitive Award.

**Competing interests** None declared.

**Patient consent for publication** Not required.

**Provenance and peer review** Not commissioned; externally peer reviewed.

**Data availability statement** Data are available on reasonable request. GWAS summary statistics are available through the GWAS catalogue. All data relevant to the study are included in the article or uploaded as supplementary information. All relevant data supporting the key findings of this study are available within the article and its supplementary information files. Other data are available from the corresponding author on reasonable request. Due to ethical and legal restrictions, individual-level data of the Rotterdam Study (RS) cannot be made publicly available. Data are available on request to the data manager of the Rotterdam Study Frank van Rooij (f.vanrooij@erasmusmc.nl) and subject to local rules and regulations. This includes submitting a proposal to the management team of RS, where upon approval, analysis needs to be done on a local server with protected access, complying with GDPR regulations.



**ORCID iDs**
Cindy Germaine Boer http://orcid.org/0000-0003-4809-0044
Rodrigo Coutinho de Almeida http://orcid.org/0000-0001-7966-056X
Jeremy Mark Wilkinson http://orcid.org/0000-0001-5577-3674
David Felson http://orcid.org/0000-0002-2668-2447

## REFERENCES

1 GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the global burden of disease study 2017. *Lancet* 2018;392:1789–858.
2 Barbour KE, Lui L-Y, Nevitt MC, *et al*. Hip osteoarthritis and the risk of all-cause and disease-specific mortality in older women: a population-based cohort study. *Arthritis Rheumatol* 2015;67:1798–805.
3 Qin J, Barbour KE, Murphy LB, *et al*. Lifetime risk of symptomatic hand osteoarthritis: the Johnston County osteoarthritis project. *Arthritis Rheumatol* 2017;69:1204–12.
4 Bijsterbosch J, Watt I, Meulenbelt I, *et al*. Clinical burden of erosive hand osteoarthritis and its relationship to nodes. *Ann Rheum Dis* 2010;69:1784–8.
5 Kwok WY, Vliet Vlieland TPM, Rosendaal FR, *et al*. Limitations in daily activities are the major determinant of reduced health-related quality of life in patients with hand osteoarthritis. *Ann Rheum Dis* 2011;70:334–6.
6 Kjeken I, Dagfinrud H, Slatkowsky-Christensen B, *et al*. Activity limitations and participation restrictions in women with hand osteoarthritis: patients' descriptions and associations between dimensions of functioning. *Ann Rheum Dis* 2005;64:1633–8.
7 Hunter DJ, Bierma-Zeinstra S. Osteoarthritis. *Lancet* 2019;393:1745–59.
8 Jiang L, Xie X, Wang Y, *et al*. Body mass index and hand osteoarthritis susceptibility: an updated meta-analysis. *Int J Rheum Dis* 2016;19:1244–54.

9 Jensen V, Bøggild H, Johansen JP. Occupational use of precision grip and forceful gripping, and arthrosis of finger joints: a literature review. *Occup Med* 1999;49:383–8.
10 Spector TD, Cicuttini F, Baker J, *et al*. Genetic influences on osteoarthritis in women: a twin study. *BMJ* 1996;312:940–3.
11 Ishimori ML, Altman RD, Cohen MJ, *et al*. Heritability patterns in hand osteoarthritis: the role of osteophytes. *Arthritis Res Ther* 2010;12:R180.
12 Zengini E, Hatzikotoulas K, Tachmazidou I, *et al*. Genome-wide analyses using UK Biobank data provide insights into the genetic architecture of osteoarthritis. *Nat Genet* 2018;50:549–58.
13 Tachmazidou I, Hatzikotoulas K, Southam L, *et al*. Identification of new therapeutic targets for osteoarthritis through genome-wide analyses of UK Biobank data. *Nat Genet* 2019;51:230–6.
14 Styrkarsdottir U, Lund SH, Thorleifsson G, *et al*. Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. *Nat Genet* 2018;50:1681–7.
15 Styrkarsdottir U, Thorleifsson G, Helgadottir HT, *et al*. Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. *Nat Genet* 2014;46:498–502.
16 den Hollander W, Boer CG, Hart DJ, *et al*. Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. *Ann Rheum Dis* 2017;76:2046–53.
17 Kloppenburg M, Kwok W-Y. Hand osteoarthritis--a heterogeneous disorder. *Nat Rev Rheumatol* 2011;8:22–31.
18 Manchia M, Cullis J, Turecki G, *et al*. The impact of phenotypic and genetic heterogeneity on results of genome wide association studies of complex diseases. *PLoS One*;8:82013.
19 Egger P, Cooper C, Hart DJ, *et al*. Patterns of joint involvement in osteoarthritis of the hand: the Chingford study. *J Rheumatol* 1995;22:1509–13.
20 Dahaghin S, Bierma-Zeinstra SMA, Ginai AZ, *et al*. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). *Ann Rheum Dis* 2005;64:682–7.
21 Castaño-Betancourt MC, Evans DS, Ramos YFM, *et al*. Novel genetic variants for cartilage thickness and hip osteoarthritis. *PLoS Genet* 2016;12:e1006260.
22 Panoutsopoulou K, Thiagarajah S, Zengini E, *et al*. Radiographic endophenotyping in hip osteoarthritis improves the precision of genetic association analysis. *Ann Rheum Dis* 2017;76:1199–206.
23 Bierma-Zeinstra SM, van Middelkoop M. Osteoarthritis: in search of phenotypes. *Nat Rev Rheumatol* 2017;13:705–6.
24 Ikram MA, Brusselle GGO, Murad SD, *et al*. The Rotterdam study: 2018 update on objectives, design and main results. *Eur J Epidemiol* 2017;32:807–50.
25 Das S, Forer L, Schönherr S, *et al*. Next-generation genotype imputation service and methods. *Nat Genet* 2016;48:1284–7.
26 Zhan X, Hu Y, Li B, *et al*. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 2016;32:1423–6.
27 Winkler TW, Day FR, Croteau-Chonka DC, *et al*. Quality control and conduct of genome-wide association meta-analyses. *Nat Protoc* 2014;9:1192–212.
28 Willer CJ, Li Y, Abecasis GR. Metal: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
29 Team RC. *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing, 2014. http://www.R-project.org/
30 Haugen IK, Englund M, Aliabadi P, *et al*. Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham osteoarthritis study. *Ann Rheum Dis* 2011;70:1581–6.
31 Kellgren JH, Lawrence JS. Radiological assessment of osteo-arthrosis. *Ann Rheum Dis* 1957;16:494–502.
32 Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res* 2016;44:D877–81.
33 Watanabe K, Taskesen E, van Bochoven A, *et al*. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017;8:1826.
34 Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, *et al*. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317–30.
35 ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature* 2012;489:57–74.
36 Wang Y, Song F, Zhang B, *et al*. The 3D genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. *Genome Biol* 2018;19:151.
37 Ward LD, Kellis M. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. *Nucleic Acids Res* 2012;40:D930–4.
38 Kalichman L, Cohen Z, Kobyliansky E, *et al*. Patterns of joint distribution in hand osteoarthritis: contribution of age, sex, and handedness. *Am J Hum Biol* 2004;16:125–34.
39 Marshall M, van der Windt D, Nicholls E, *et al*. Radiographic hand osteoarthritis: patterns and associations with hand pain and function in a community-dwelling sample. *Osteoarthritis Cartilage* 2009;17:1440–7.
40 den Hollander W, Pulyakhina I, Boer C, *et al*. Annotating transcriptional effects of genetic variants in disease-relevant tissue: transcriptome-wide allelic imbalance in osteoarthritic cartilage. *Arthritis Rheumatol* 2019;71:561–70.
41 Steinberg J, Ritchie GRS, Roumeliotis TI, *et al*. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. *Sci Rep* 2017;7:8935.
42 Reynard LN, Loughlin J. Insights from human genetic studies into the pathways involved in osteoarthritis. *Nat Rev Rheumatol* 2013;9:573–83.
43 Richard D, Liu Z, Cao J, *et al*. Evolutionary selection and constraint on human knee chondrocyte regulation impacts osteoarthritis risk. *Cell* 2020;181:362–81.
44 Vaes RBA, Rivadeneira F, Kerkhof JM, *et al*. Genetic variation in the GDF5 region is associated with osteoarthritis, height, hip axis length and fracture risk: the Rotterdam study. *Ann Rheum Dis* 2009;68:1754–60.
45 Hämäläinen S, Solovieva S, Vehmas T, *et al*. Genetic influences on hand osteoarthritis in Finnish women--a replication study of candidate genes. *PLoS One* 2014;9:e97417.
46 Phillips JE, Corces VG. CTCF: master weaver of the genome. *Cell* 2009;137:1194–211.
47 Später D, Hill TP, O'sullivan RJ, *et al*. Wnt9a signaling is required for joint integrity and regulation of Ihh during chondrogenesis. *Development* 2006;133:3039–49.
48 Hartmann C, Tabin CJ. Wnt-14 plays a pivotal role in inducing synovial joint formation in the developing appendicular skeleton. *Cell* 2001;104:341–51.
49 Maupin KA, Droscha CJ, Williams BO. A comprehensive overview of skeletal phenotypes associated with alterations in Wnt/β-catenin signaling in humans and mice. *Bone Res* 2013;1:27–71.
50 Zhou Y, Wang T, Hamilton JL, *et al*. Wnt/β-catenin Signaling in Osteoarthritis and in Other Forms of Arthritis. *Curr Rheumatol Rep* 2017;19:53.
51 Lietman C, Wu B, Lechner S, *et al*. Inhibition of Wnt/β-catenin signaling ameliorates osteoarthritis in a murine model of experimental osteoarthritis. *JCI Insight* 2018;3. doi:10.1172/jci.insight.96308. [Epub ahead of print: 08 Feb 2018].
52 Nelson MR, Tipney H, Painter JL, *et al*. The support of human genetic evidence for approved drug indications. *Nat Genet* 2015;47:856–60.
53 Mars N, Koskela JT, Ripatti P, *et al*. Polygenic and clinical risk scores and their impact on age at onset and prediction of cardiometabolic diseases and common cancers. *Nat Med* 2020;26:549–57.

# Supplementary Text

Supplementary materials for *Boer C.G., et al., Stratified hand phenotypes identifies WNT9A as novel gene associated with thumb osteoarthritis.*

**Contents**

## Supplementary methods

*Description of the Rotterdam Study (RS)*

The Rotterdam Study(RS) is a prospective population based cohort consisting of elderly inhabitants, 45 years and older, of the Omnoord district in the city of Rotterdam, the Netherlands[1]. The RS has been ongoing since 1990 to study the determinants of chronic disabling disease in the elderly. The Rotterdam Study I (RS-I) is the first cohort, of 7,983 persons living in the Ommoord district of Rotterdam in the Netherlands. All subjects were aged 55 years and older, recruitment of participants started in 1990. The Rotterdam Study II (RS-II) started in 1999 when 3,011 participants moved into the study since they became 55 years of age or moved into the study district. The Rotterdam Study III (RS-III) started in 2006 with all 3,932 participants aged 45 years and older from the study district not yet included in the study. The present study includes all participants for whom radiographs of the hand joints at baseline visit were present. During the visit to the study center (online supplementary Table S1), bilateral hand radiographs were made during the baseline visit, which were examined and scored by trained radiologists[2]. For both hands all distal interphalangeal joints (DIP), interphalangeal joints (PIP), metacarpophalangeal joints (MCP), thumb interphalangeal joint(IP), first carpometacarpal (CMC1) joint and the trapezioscaphoid (TS) joint were scored according to the Kellgren-Lawrence (KL) OA severity grading scale[3]. . For 1,609 participants, we were unable to score one or more joints, or genetic data was ,not available, which left us with a total of 8,691 participants to perform the study. The interobserver reliability for KL≥2 was: DIP: κ=0.60, PIP =0.61, MCP =0.63 and CMC1/TS=0.75[2].

The Rotterdam Study has been approved by the Medical Ethics Committee of the Erasmus MC (registration number MEC 02.1015) and by the Dutch Ministry of Health, Welfare and Sport (Population Screening Act WBO, license number 1071272-159521-PG). The Rotterdam Study has been entered into the Netherlands National Trial Register (NTR; www.trialregister.nl) and into the WHO International Clinical Trials Registry Platform (ICTRP; www.who.int/ictrp/network/primary/en/) under shared catalogue number NTR6831. All participants provided written informed consent to participate in the study and to have their information obtained from treating physicians

*Description of the Framingham Heart Study(FHS)*

The original Framingham Study was a population-based sample of adults (ages 28-61 years) that began in 1948[4] . The Framingham Offspring Study is composed of children of the original Framingham Heart Study participants, and the children's spouses[5]. As part of an ancillary study in 1992 to 1995, Offspring (and their spouses) were contacted by mail and telephone call to participate in a visit to assess hand OA. About 1,800 individuals (ages 28-82 years) were examined, representing about 65% of those contacted. Of these individuals, 1,293 participants returned for another hand examination in 2002 to 2005[6]. As osteoarthritis often does not present until a later date, we have used data from the 2002 to 2005 visit. Of which 1,203 had genotyping data available for analysis. Individuals underwent bilateral poster-anterior hand radiographs, which were read by a trained musculoskeletal radiologist. The bilateral 2nd-5th DIP, 2nd-5th PIP, 1st-5th MCP, IP, thumb base (carpometacarpal) joint and wrist joints were scored according to KL-score with good inter-reader reliability (weighted $\kappa$=0.76).

*Patient and Public Involvement*

Patients were involved in the design of this study, through the Dutch Arthritis Association (DAA)through the founding of this research (DAA 2010_017). Patients and the general Public will be informed of the results through the dedicated website of the Dutch Arthritis Association (https://reumanederland.nl/), and via the Erasmus MC Rotterdam Osteoarthritis Research (ROAR) twitter account (@roar_NL).

*GWAS, discovery, replication and meta-analysis*

Genome-wide association(GWAS) methods of the discovery cohort have been described previously[7], briefly genotyped variants were imputed after quality control using the michigain imputation sever(HRC panel v.1.1[8]). Genetic dosages were used to investigate association with the stratified hand OA phenotypes using RVTESTS[9]. All performed GWAS, including replication analysis, were adjusted for age, sex and the first four genetic principal components. We did not include BMI in our model, as inclusion of heritable and causally associated covariates can introduce "collider bias"[10]. Variants were considered for replication if p-value$\leq 1*10^{-06}$. Meta-analysis between discovery (RS) and replication (FHS) was performed using inverse variance weighting (METAL[11]). Variants were considered "replicated and genome-wide significant if their replication p-value<0.05, had

the same direction of the beta and the meta-analysis p-value<$5*10^{-08}$ (Genome-wide significance threshold)[12]. Variants were considered genome-wide suggestive when they were replicated and the meta-analysis p-value<$1*10^{-06}$. Independence for each signal was determined by conditional-joint analysis(Jo-Co) in GCTA[13]. Manhattan, QQ-plots and heatmap plots were made in R[14] using the CRAN software packages qqman, gplots and RcolorBrewer. Images were saved in. eps format, font size, style and additional text were added/modified using Adobe Illustrator.

*Hand osteoarthritis Cluster Analysis*

Cluster analysis was performed on all hand joints in the RS cohorts (n=8,691). The goal was to organize the observed data into meaningful clusters using hierarchical clustering of a Euclidean distance matrix using Ward's method. For each radiographic measurement separately (joint space narrowing, osteophyts, and KL-score) we performed normalization of the data through scaling the data across the joints and calculated a distance matrix based on Euclidean distances. Next, we used Ward's agglomerative hierarchical clustering method to generate tree diagrams. Where the vertical axis denoted the linkage distance. We used multidimensional scaling (MDS) to further detect biologically interpretable clusters between the joints groups. For the multidimensional scaling we used 2 dimensions. We scaled the data and calculated a distance matrix based on Euclidean distances to be used in the MDS. These cluster analysis were also performed on the radiographic KLscore of each measured joint of the hand separately, without grouping joints per type. Clusters were determined by comparing the results from all cluster analysis. The following clusters were recognized: finger KLsum included all DIP and PIP joints, excluding the IP joints, as these cluster either more with the MCP in the KL-grade, or the DID/PIP dependent on the radiographic feature examined (online supplementary Figure S1 and S2). The Finger KLsum score includes in total the KL grade of 16 joints and can range from 0 to 64, where a score of 64 means that the maximum KL grade (4) was assigned to all joint included in the KLsum score. The thumb KLsum included the TS and the CMC1 joints, the IP joint was excluded as this did not cluster with CMC1 or TS joints in any of the radiographic features examined. The Thumb KLsum score includes in total the KL grade of 4 joints can range from 0 to 16, where a score of 16 means that the maximum KL grade (4) was assigned to all joints included in the KLsum score. The hand KLsum included all DIP, PIP, MCP, IP and CMC1 joints, the IP and TS joint

were excluded, as these do not consistently cluster with the rest of the joints in the radiographic features examined. The Hand KLsum score includes in total the KL grade of 30 joints and can range from 0 to 120, where a score of 120 means that the maximum KL grade (4) was assigned to all joints included in the KLsum score. All cluster analysis were performed in R.

*Lookup in DECODE and UKbiobank osteoarthritis GWAS*

The SNVs identified in the discovery GWAS (RS) and replicated in the replication cohort (FHS) were also examined for association with clinical osteoarthritis in a meta-analysis of the Icelandic DECODE population cohort and the united kingdom based UKbiobank population cohort[15,16]. Information on osteoarthritis was derived from a national Icelandic hip or knee arthroplasty registry, electronic health records (using ICD10 codes), and a dedicated hand osteoarthritis database[15,16].

*Variant functional annotation*

A locus was defined as the region 500kb upstream and 500kb downstream from the lead SNP. For each lead variant SNPs in high LD ($r^2 \geq 0.8$) were determined and annotated using annotation provided by FUMA and HaploregV4. 1[17,18]. All variants were and gene regulatory region annotation were provided by the SNP2GENE tool from FUMA, Haploreg V4 annotation and from the ROADMAP and ENCODE projects[19,20]. Intersection of the variant with gene regulatory elements as predicted by histone post-translational modifications, were made by the ROADMAP project[19]. CTCF-protein binding Chip-seq peaks in primary osteoblast cells were generated by ENCODE[20] and visualized via UCSC genome-browser[21]. Annotation of the variant location, number of proteins bound and transcription factor(TF) binding motifs change was done via Haploreg V1.4. as described previously[18]. TF binding to gene promoter locations were taken from the ENCODE Transcription Factor Binding Site Profiles dataset[20] accessed through harmonizome[22].

*Human embryonic cartilage ATAC-seq*

Intersection of SNVs with open chromatin regions in human embryonic cartilage was done using ATAC-seq data. For a detailed description of this human embryonic ATAC seq dataset see[23]. We have acquired chromatin accessibility from human embryonic cartilage ATAC-seq datasets at E59 of gestation[23] to investigate if our lead SNV and variants in high

LD co-localized with these open chromatin regions. All variants in LD with the lead variant (hg19 coordinates) were intersected with E59 ATAC-seq peaks from four cartilage tissues (proximal femur, distal femur, proximal tibia and distal tibia) using the UCSC Genome Browser Table Browser tool. ATAC-seq peaks from these tissues were also used to map regulatory elements in the Wnt locus on human chromosome 1.

*Gene prioritization*

Candidate genes in the locus were defined as: all genes annotated to be (partially) located within the locus (500 kb upstream and downstream of the lead SNV).All genes fitting this definition were considered as potential causal gene, in total we analyzed 18 genes (supplementary table 3). Genes were prioritized based on several lines of evidence: 1)eQTL-analysis consisted of two separate hip OA cartilage datasets (n=29 and n=87) of which genotypes and RNA-seq data were available[24,25]. For each dataset an eQTL analysis was performed for each lead SNP with all genes in the locus. To increase detection power, we then meta-analyzed the result from both datasets together using a weighted meta-analysis based on p-values, sample size and direction of effect in METAL[11]. 3)The 3D chromatin structure of the locus was examined using Capture Hi-C data from human mesenchymal stem cells (hMSCs)[26].We visualized the chromatin interactions between the lead SNV and possible causal SNV(s) with promoter regions of *WNT9A/WNT3A/JMJD4* and *SNAP47* and the rest of the investigated locus.

*RNA-sequencing for differential expression*

Differential gene expression between OA lesioned and Preserved cartilage: Post-RNA isolation (Qiagen RNeasy Mini Kit, RIN >7) of 40 knee (15 paired preserved (P) and OA lesioned (OAL), 7 P only and 3 OAL only) and 28 hip (six paired P and OAL, 14 P only and 2 OAL only) cartilage samples (supplementary table 3), paired-end 2×100 bp RNA library sequencing (Illumina TruSeq RNA-Library Prep Kit, Illumina HiSeq2000) resulted in an average of 10 million fragments per sample. Reads were aligned using GSNAP against GRCh37/hg19, in which SNPs from the Genome of the Netherlands consortium with a minor allele frequency (MAF) >1% were masked to prevent alignment bias. Number of

fragments per gene were used to assess quantile-adjusted conditional maximum likelihood (edgeR, R-package)[27]. Subsequently, differential gene expression analysis was performed pairwise between P and OAL samples for which we had RNA of both (n=21).

*Expression quantitative trait loci analysis*

The eQTL-analysis consisted of two separate OA cartilage datasets (n=29 and n=87) of which genotypes and RNA-seq data were available. RNA-expression and eQTL analysis of the first dataset of 29 samples has been previously described here[28]**.** The second dataset consisting of 87 samples were collected and RNA was extracted as previously described[25]. Post RNA isolation, multiplexed libraries were sequenced on the Illumina HiSeq 2000 (75bp paired-end read length). Sample QC was carried out using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/) and transcript-level quantification was performed using salmon[29] based on the GRCh38 cDNA assembly [http://ftp.ensembl.org/pub/release-87/fasta/homo_sapiens/cdna/]. Transcript-level estimates were summarized to gene-level estimates (scaled transcripts per million) based on Ensembl gene IDs using tximport[30]. Only genes with ≥1 count per million in ≥20% samples were kept, with 87 low-grade cartilage samples and 15,249 genes post QC. All 87 individuals were genotyped using Illumina HumanCoreExome. Genotypes were called using GenCall and mapped to GRC37/hg19. Following sample and variant QC, we imputed up to HRC panel v1.1 using the Michigan imputation server (https://imputationserver.sph.umich.edu/index.html). We followed the GTEx approach for eQTL analysis[31]. Briefly, we normalized gene expression between samples using weighted trimmed mean of M-values implemented in edgeR[27]. For each gene, expression across samples was normalized using an inverse normal transformation. To identify cis-eQTLs within 1Mb either direction of a gene transcription start site, we used the GTEx modified version of FastQTL (https://github.com/francois-a/fastqtl; v6p), including 15 Probabilistic Estimation of Expression Residuals (PEER) factors[32], sex and genotype array as covariates. We generated empirical p-values, with a 5% Storey-Tibshirani FDR cut-off to identify genes with a significant eQTL[33]. The normalised effect size (NES) of each eQTL is reported for the alternate allele.

*Methylation quantitative trait loci analysis*

We used cartilage CpG methylation and genotype data that had been generated previously using Illumina's Infinium HumanMethylation450 array and HumanOmniExpress array, respectively[34]. Methylation and genotype data were generated from 87 patients who had undergone knee or hip joint arthroplasty: 57 knee OA patients, 14 hip OA patients and 16 control patients who had undergone hip replacement due to a neck-of-femur (NOF) fracture. If the SNP reported as associated with OA in the GWAS was directly genotyped on the HumanOmniExpress array, that SNP data was used by us. If the SNP was not, we searched for and, where possible, used a proxy SNP that was in perfect or high LD (pairwise r2>0.7) with the association SNP. Proxies were derived from a candidate list using LDlink's LDproxy tool[35] and European population data. Where multiple proxies were identified, the one with the highest r2 relative to the association SNP was chosen. For each locus, we covered a 1Mb region encompassing 500kb upstream and 500kb downstream of the association SNP. For each CpG within the 1Mb, linear regression was used to measure the relationship between methylation in the form of M-values and genotype (0, 1 or 2 copies of the minor allele) at the OA association SNP or its proxy. Age, sex and joint site/condition were added into the model as covariates. Methylation status is reported using β-values (ranging from 0 for no methylation to 1 for 100% methylation). mQTL calculations were performed using Matrix eQTL[36] implementing a false discovery rate (FDR) estimation that is based on the Benjamini–Hochberg FDR procedure[37] and which accounts for the number of tests performed.

# References


1. Ikram MA, Brusselle GGO, Murad SD, et al. The Rotterdam Study: 2018 update on objectives, design and main results. Eur J Epidemiol 2017;32:807-50.
2. Dahaghin S, Bierma-Zeinstra SM, Ginai AZ, Pols HA, Hazes JM, Koes BW. Prevalence and pattern of radiographic hand osteoarthritis and association with pain and disability (the Rotterdam study). Ann Rheum Dis 2005;64:682-7.
3. Kellgren JH, Lawrence JS. Radiological Assessment of Osteo-Arthrosis. Ann Rheum Dis 1957;16:494-502.
4. Dawber TR, Meadors GF, Moore FE. Epidemiological Approaches to Heart Disease: The Framingham Study *. Am J Public Health Nations Health 1951;41:279-86.
5. Kannel WB, Feinleib M, McNamara PM, Garrison RJ, Castelli WP. An investigation of coronary heart disease in families. The Framingham offspring study. Am J Epidemiol 1979;110:281-90.
6. Haugen IK, Englund M, Aliabadi P, et al. Prevalence, incidence and progression of hand osteoarthritis in the general population: the Framingham Osteoarthritis Study. Ann Rheum Dis 2011;70:1581-6.
7. den Hollander W, Boer CG, Hart DJ, et al. Genome-wide association and functional studies identify a role for matrix Gla protein in osteoarthritis of the hand. Ann Rheum Dis 2017;76:2046-53.
8. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48:1279-83.
9. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics 2016;32:1423-6.
10. Day FR, Loh PR, Scott RA, Ong KK, Perry JR. A Robust Example of Collider Bias in a Genetic Association Study. Am J Hum Genet 2016;98:392-3.
11. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26:2190-1.
12. Pe'er I, Yelensky R, Altshuler D, Daly MJ. Estimation of the multiple testing burden for genomewide association studies of nearly all common variants. Genetic epidemiology 2008;32:381-5.
13. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am J Hum Genet 2011;88:76-82.
14. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria., 2014. at http://www.R-project.org/.)
15. Styrkarsdottir U, Lund SH, Thorleifsson G, et al. Meta-analysis of Icelandic and UK data sets identifies missense variants in SMO, IL11, COL11A1 and 13 more new loci associated with osteoarthritis. Nat Genet. United States2018:1681-7.
16. Styrkarsdottir U, Thorleifsson G, Helgadottir HT, et al. Severe osteoarthritis of the hand associates with common variants within the ALDH1A2 gene and with rare variants at 1p31. Nat Genet 2014;46:498-502.
17. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun 2017;8:1826.
18. Ward LD, Kellis M. HaploReg v4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. Nucleic Acids Res 2016;44:D877-81.
19. Kundaje A, Meuleman W, Ernst J, et al. Integrative analysis of 111 reference human epigenomes. Nature 2015;518:317-30.
20. An integrated encyclopedia of DNA elements in the human genome. Nature 2012;489:57-74.
21. Casper J, Zweig AS, Villarreal C, et al. The UCSC Genome Browser database: 2018 update. Nucleic Acids Res 2018;46:D762-d9.

22. Rouillard AD, Gundersen GW, Fernandez NF, et al. The harmonizome: a collection of processed datasets gathered to serve and mine knowledge about genes and proteins. Database (Oxford) 2016;2016.
23. Richard D, Liu Z, Cao J, et al. Evolutionary Selection and Constraint on Human Knee Chondrocyte Regulation Impacts Osteoarthritis Risk. Cell 2020.
24. Steinberg JS, S.; Butterfield, N.C.; Roumeliotis,T.I.; Fontalis A.; Clark, M.J.; Jayasuriya, R.L.; Swift,D.; Shah,K.M.; Curry,K.F.; Brooks,R.A.; McCaskie, A.W.; Lelliott, C.J.;Choudhary, J.S.; Bassett, J.H.D.; Williams,G.R.; Wilkinson, J.M.; Zeggini, E. Decoding the genomic basis of osteoarthritis. bioRxiv 8358502019.
25. Steinberg J, Ritchie GRS, Roumeliotis TI, et al. Integrative epigenomics, transcriptomics and proteomics of patient chondrocytes reveal genes and pathways involved in osteoarthritis. Sci Rep 2017;7:8935.
26. Wang Y, Song F, Zhang B, et al. The 3D Genome Browser: a web-based browser for visualizing 3D genome organization and long-range chromatin interactions. Genome Biol 2018;19:151.
27. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 2010;26:139-40.
28. den Hollander W, Pulyakhina I, Boer C, et al. Annotating Transcriptional Effects of Genetic Variants in Disease-Relevant Tissue: Transcriptome-Wide Allelic Imbalance in Osteoarthritic Cartilage. Arthritis Rheumatol2019:561-70.
29. Patro R, Duggal G, Love MI, Irizarry RA, Kingsford C. Salmon provides fast and bias-aware quantification of transcript expression. Nat Methods 2017;14:417-9.
30. Soneson C, Love MI, Robinson MD. Differential analyses for RNA-seq: transcript-level estimates improve gene-level inferences. F1000Res 2015;4:1521.
31. The Genotype-Tissue Expression (GTEx) project. Nat Genet 2013;45:580-5.
32. Stegle O, Parts L, Piipari M, Winn J, Durbin R. Using probabilistic estimation of expression residuals (PEER) to obtain increased power and interpretability of gene expression analyses. Nat Protoc 2012;7:500-7.
33. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440-5.
34. Rushton MD, Reynard LN, Young DA, et al. Methylation quantitative trait locus analysis of osteoarthritis links epigenetics with genetic risk. Hum Mol Genet 2015;24:7432-44.
35. Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015;31:3555-7.
36. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics 2012;28:1353-8.
37. Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 2001;125:279-84.

# Supplementary Tables

Supplementary materials for *Boer C.G., et al., Stratified hand phenotypes identifies WNT9A as novel gene associated with thumb osteoarthritis.*

**Contents**

**Supplementary tables**

## Supplementary Tables

**Supplementary table 1: human cartilage eQTL summary statistics meta-analysis reults.**

| | Meta-Analysis | | | | Heterogeneity analysis | | | | |
|---|---|---|---|---|---|---|---|---|---|
| **Gene** | **Weight (n)** | **Zscore** | **P-value** | **Direction** | **HetISq** | **ChiSq** | **Df** | **P-value** | **Weight (n)** |
| ARF1 | 116 | -1.81 | 0.0703 | -- | 0 | 0.042 | 1 | 0.8373 | 116 |
| C1orf35 | 116 | 1.818 | 0.06901 | ++ | 0 | 0.223 | 1 | 0.6369 | 116 |
| CDC42BPA | 116 | -0.285 | 0.776 | -- | 0 | 0 | 1 | 0.9968 | 116 |
| GJC2 | 29 | 1.192 | 0.2334 | ?+ | 0 | 0 | 0 | 1 | 29 |
| GUK1 | 116 | 0.229 | 0.8192 | +- | 0 | 0.215 | 1 | 0.6428 | 116 |
| IBA57 | 116 | 2.01 | 0.04439 | ++ | 0 | 0.81 | 1 | 0.3682 | 116 |
| JMJD4 | 29 | -0.686 | 0.4928 | ?- | 0 | 0 | 0 | 1 | 29 |
| LINC01641 | . | . | . | . | . | . | . | . | . |
| MIR3629 | . | . | . | . | . | . | . | . | . |
| MIR5008 | 29 | 0.021 | 0.9832 | ?+ | 0 | 0 | 0 | 1 | 29 |
| MRPL55 | 116 | 0.105 | 0.9165 | ++ | 0 | 0.001 | 1 | 0.9727 | 116 |
| OBSCN | 116 | 0.412 | 0.6802 | ++ | 0 | 0.015 | 1 | 0.9021 | 116 |
| OBSCN-AS1 | . | . | . | . | . | . | . | . | . |
| PRSS38 | 29 | 0.135 | 0.8925 | ?+ | 0 | 0 | 0 | 1 | 29 |
| SNAP47 | 29 | -1.569 | 0.1166 | ?- | 0 | 0 | 0 | 1 | 29 |
| WNT3A | 29 | -0.123 | 0.9024 | ?- | 0 | 0 | 0 | 1 | 29 |
| WNT9A | 116 | 1.273 | 0.2029 | +- | 81.8 | 5.489 | 1 | 0.01914 | 116 |
| ZNF678 | 116 | 1.605 | 0.1085 | ++ | 0 | 0.177 | 1 | 0.6738 | 116 |

OA: osteoarthritis, SE: standard Error, METAL implements Cochran's Q-test for heterogeneity: Df: degrees of Freedom, ChiSq: Chi-Squared test result of heterogeneity, HetISq: heterogeneity $I^2$ statistic.

**Supplementary table 2: Differential expression in OA cartilage**

**Differential expression OA preserved vs. OA lesioned cartilage (n=21)**

| GeneNames | baseMean | FoldChange | log2FoldChange | SE | P-value | P-value adjusted (FDR) |
|---|---|---|---|---|---|---|
| ARF1 | 1413.89 | 1.00 | $5.7*10^{-03}$ | $3.7*10^{-02}$ | 0.88 | 0.95 |
| C1orf35 | 57.34 | 0.95 | $-6.7*10^{-02}$ | 0.13 | 0.62 | 0.82 |
| CDC42BPA | 579.25 | 1.06 | $8.9*10^{-02}$ | $9.8*10^{-02}$ | 0.36 | 0.63 |
| GJC2 | 6.57 | 0.67 | -0.58 | 0.23 | $1.0*10^{-02}$ | $7.4*10^{-02}$ |
| GJC2 | 16.45 | 0.78 | -0.37 | 0.23 | $8.6*10^{-02}$ | 0.53 |
| GUK1 | 697.69 | 1.02 | $2.6*10^{-02}$ | 0.10 | 0.80 | 0.92 |
| IBA57 | 58.63 | 0.84 | -0.25 | 0.13 | $4.7*10^{-02}$ | 0.20 |
| JMJD4 | 129.36 | 0.97 | $-4.7*10^{-02}$ | 0.11 | 0.67 | 0.84 |
| LINC01641 | 0 | 1 | 1.65 | 2.93 | 1 | 1 |
| MIR5008 | 0.10 | 1.51 | 0.60 | 2.46 | 0.34 | 0.87 |
| MRPL55 | 113.24 | 1.04 | 0.057 | 0.13 | 0.65 | 0.84 |
| OBSCN | 376.52 | 0.60 | -0.74 | 0.18 | $2,7*10^{-05}$ | $7.8*10^{-04}$ |
| OBSCN-AS1 | 8.94 | 0.55 | -0.86 | 0.23 | $1.8*10^{-04}$ | $3.6*10^{-03}$ |
| PRSS38 | 0.33 | 0.79 | -0,34 | 2.92 | 0.64 | 1 |
| SNAP47 | 257.91 | 1.11 | 0.16 | $8.6*10^{-02}$ | $7.0*10^{-02}$ | 0.26 |
| **WNT9A** | **98.48** | **2.42** | **1.27** | **0.20** | $\mathbf{4.9*10^{-10}}$ | $\mathbf{9.4*10^{-08}}$ |
| ZNF678 | 38.41 | 1.08 | 0.11 | 0.18 | 0.54 | 0.77 |

OA: osteoarthritis, SE: standard Error, BaseMean: mean expression gene (lesioned/preserved cartilage).

**Supplementary Table 3: Transcription Factor binding motifs affected by rs11588850**

| **SNV** | **Chr** | **Ref** | **Alt** | | | | | |
|---|---|---|---|---|---|---|---|---|
| rs11588850 | 1 | A | G | | | | | |
| | | | | | | | | |
| **Transcription Factor** | **Motif Altered** | **Ref** | **Alt** | **Strand** | ***WNT9A*** | ***WNT3A*** | ***SNAP47*** | ***JMJD4*** |
| RAD21 | Rad21_disc8 | 9.8 | 11.2 | - | A549 ECC-1 GM12878 H1-hESC HCT116 HCT116 HeLa-S3 HepG2 IMR-90 K562 MCF-7 SK-N-SH | H1-hESC | . | . |
| AhR/ARNT1 complex | AhR::Arnt_1 | 10.4 | 15 | - | . | . | . | . |
| ARNT1 | Arnt_1 | -14.7 | -10.1 | + | . | . | . | . |

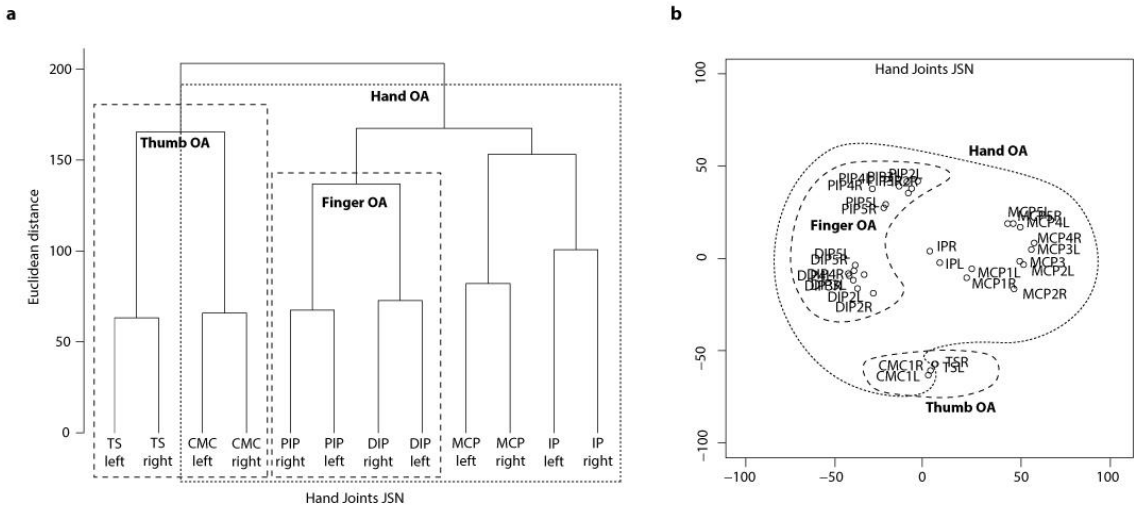
# Supplementary Figures

Supplementary materials for *Boer C.G., et al., Stratified hand phenotypes identifies WNT9A as novel gene associated with thumb osteoarthritis.*

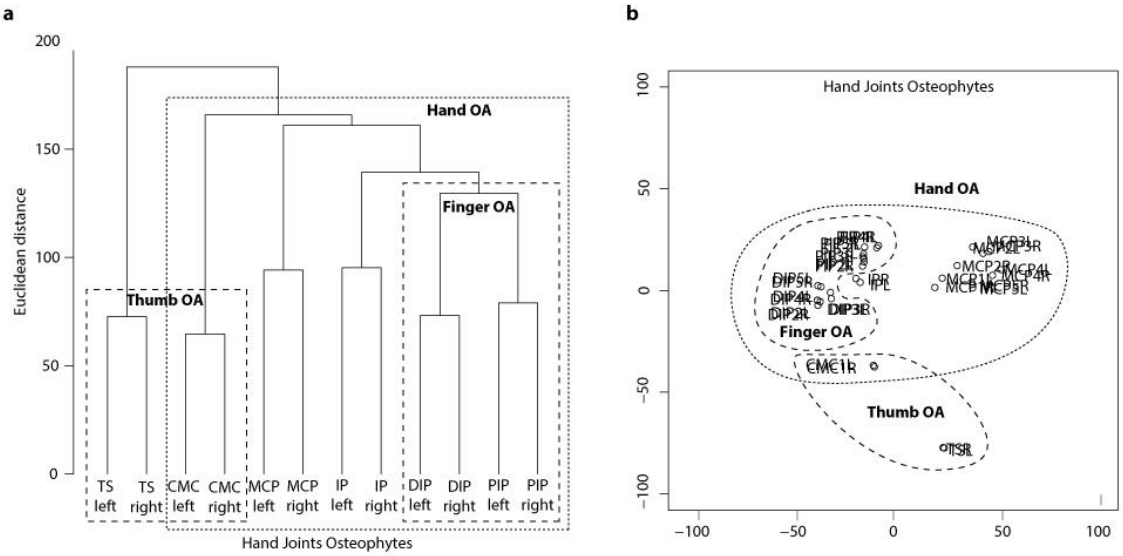
**Contents**

**Supplementary figures**

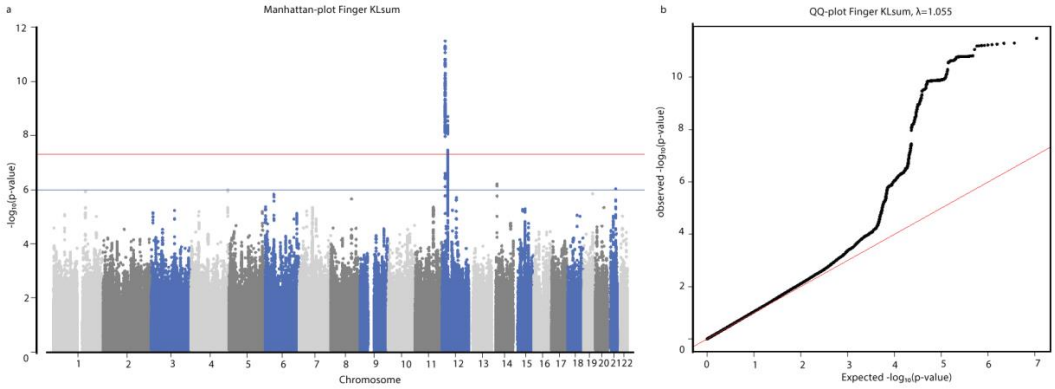
## Supplementary Figures



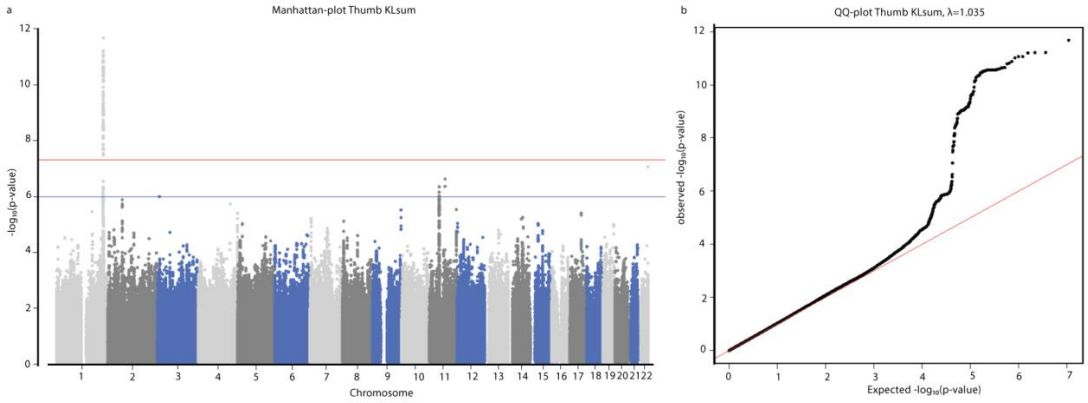

**online supplementary Figure S1: Tree-dendrogram and multidimensional scaling (MDS) plot of JSN scores in the joints of the hand**. a) tree-dendogram of complete hierarchal clustering of Euclidean distance matrix of JSN scores of groups of joints. Joints were grouped per 'type' and per hand, left or right. b) MDS plot of JSN scores for all hand joints left (L) and right (R), not grouped per "type". Data consisted of all available data from RSI, RSII and RSIII(n=8,691)**.** Selected clusters are depicted by the different dashed lines. OA: osteoarthritis, JSN: joint space narrowing, TS: trapezioscaphoid joints, CMC: carpometacarpal joints, PIP: proximal interphalangeal joints, DIP: distal interphalangeal joints, MCP: metacarpophalangeal joints and IP: interphalangeal joints. L/R: left or right joint, number denotes which joint, i.e., PIP2L: the second PIP joint at the left hand.



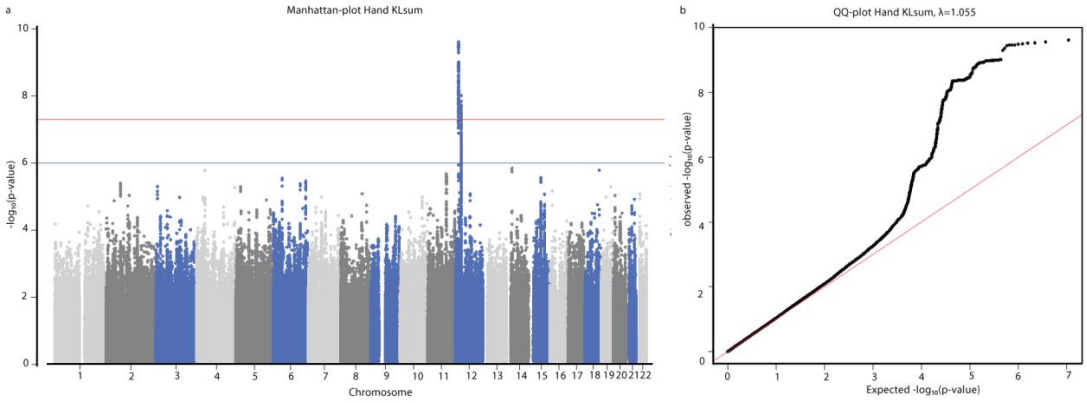

**online supplementary Figure S2: Tree-dendrogram and multidimensional scaling (MDS) plot of OST scores in the joints of the hand**. a) tree-dendogram of complete hierarchal clustering of Euclidean distance matrix of OST scores of groups of joints. Joints were grouped per 'type' and per hand, left or right. b) MDS plot of OST scores for all hand joints left (L) and right (R), not grouped per "type". Data consisted of all available data from RSI, RSII and RSIII (n=8,691)**.** Selected clusters are depicted by the different dashed lines. OA: osteoarthritis, OST: Osteophytosis, TS: trapezioscaphoid joints, CMC: carpometacarpal joints, PIP: proximal interphalangeal joints, DIP: distal interphalangeal joints, MCP: metacarpophalangeal joints and IP: interphalangeal joints. L/R: left or right joint, number denotes which joint, i.e., PIP2L: the second PIP joint at the left hand.




**Supplementary figure 3: Manhattan plot and QQ plot for meta-analysis results of Finger KLsum**.
Meta-analysis results are from RS-I, RS-II and RS-III cohorts.




**Supplementary figure 4: Manhattan plot and QQ plot for meta-analysis results of Thumb KLsum**. Meta-analysis results are from RS-I, RS-II and RS-III cohorts.




**Supplementary figure 5: Manhattan plot and QQ plot for meta-analysis results of Hand KLsum**. Meta-analysis results are from RS-I, RS-II and RS-III cohorts.